IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

ANALYSIS OF PREDICTIVE ANALYTICS USING MACHINE LEARNING ALGORITHMS IN HEALTH CARE SYSTEM

[1]Mrs.T.Charanya Nagammal, [2]Dr.K.Chitra
[1]Research Scholar, [2]Assistant Professor,
[1] Department of Computer Science,
[1] Madurai Kamaraj University, Madurai, Tamilnadu, India

Abstract: Predictive Analytics is a technique used to determine the future performance based on the current data and historical data. In the healthcare industry, predictive analytics is a tough challenge. Data analytics aids health clinicians in predicting various diseases by allowing them to make predictions based on a larger quantity of data in order to decrease the death rate. Using machine learning algorithms in health care can help us to find patterns and insights in medical data that would be impossible to find manually. The integration of machine learning and medicine is primarily aimed at enhancing the efficiency, accuracy, and personalization of healthcare. There is more number of machine learning algorithms are used for the process of predictive analytics. Some approaches are Artificial Neural Network (ANN), Decision tree, Fuzzy Logic, K-Nearest Neighbour (KNN), Naïve Bayes and Support Vector Machine (SVM)….etc. This paper provides instill of the existing algorithms and it gives an overall summary of the existing work.

Keywords: Predictive Analytics, Machine Learning, Healthcare

I. INTRODUCTION:

There is an almost universal definition shared with proponents of the ideology of big data, is that "Big Data sets a situation in which data sets have increased at such huge sizes that conventional technologies of information, can no longer manage them effectively, either the size or the extent and the growth of the data set. The world has become submerged by a large amount of data. Every moment is equivalent to the generation of thousands data.

As Big Data technology evolves, businesses are turning to Predictive Analytics to help them enhance consumer engagement, streamline operations, and cut operational costs. The combination of real-time Big Data streams with Predictive Analytics— also known as “never-ending processing”— has the potential to give a significant competitive advantage. The influence of data mining is more on analysis of healthcare industry data. This enables the health systems to use and analyze the data to identify best practices to improve health care and reduce clinical and investigation costs. Data mining applications do not make predicting the outcome of an illness simple. Historical cases stored within the datasets can be used to build the best models with more accuracy to predict the outcome of a disease. Every technique in data mining facilitates a different objective based on the modeling for medical data analysis. Classification and predictive analytics are the most used common modeling techniques. Classification is commonly used to forecast outcomes on categorical labels and the predictive technique is applied for continuously evaluated features.

II. PREDICTIVE ANALYTICS IN HEALTH CARE:

Predictive analytics in healthcare involves analyzing historical data to identify patterns and predict future outcomes, such as disease onset, patient outcomes, or resource needs. It can help in early disease detection, personalized treatment plans, resource allocation, and overall efficiency improvement in healthcare delivery. It is highly useful in disease prediction in several ways:

2.1Early Detection:

By analyzing large datasets containing patient information, predictive analytics can identify patterns and risk factors associated with the onset of various diseases. Early detection allows for timely interventions, potentially preventing the progression of the disease or minimizing its impact.

2.2 Risk Stratification:

Predictive models can stratify individuals based on their likelihood of developing certain diseases. This enables healthcare providers to prioritize resources and interventions for high-risk individuals, such as targeted screening programs or preventive interventions.

2.3 Personalized Risk Assessment:

Predictive analytics takes into account individual characteristics, including demographics, medical history, genetic predisposition, lifestyle factors, and environmental exposures. This personalized approach allows for more accurate risk assessments tailored to each individual's profile. By identifying individuals at high risk of developing specific diseases, healthcare preventive interventions providers can implement preventive interventions, such as lifestyle modifications, vaccinations, or Screening Tests, To Mitigate Risk Factors And Prevent Disease Onset.

2.4 Population Health Management:

Predictive analytics can inform population-level health management strategies by identifying trends and patterns in disease prevalence, distribution, and risk factors within specific populations. This information enables healthcare organizations to allocate resources effectively and develop targeted public health interventions.

2.5 Research and Development:

Predictive analytics can contribute to medical research by identifying novel risk factors, biomarkers, or genetic predispositions associated with certain diseases. This information can inform the development of new diagnostic tools, treatment modalities, or preventive measures.

Overall, predictive analytics in disease prediction empowers healthcare providers, policymakers, and individuals to take proactive steps towards preventing disease, improving population health outcomes, and reducing healthcare costs

III. MACHINE LEARNING ALGORITHMS IN HEALTH CARE:

Data analytics mechanism experimented in diverse healthcare scenarios, towards developing the most efficient machine learning algorithms to be used depending on the healthcare scenario's requirements and datasets, for efficiently predicting the onset of a disease. In general, Machine Learning tools have grown in popularity in the healthcare area during the last few decades, where a variety of ML algorithms, including K-Nearest Neighbors(KNN), Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), Neural Networks (NN), Naïve Bayes, and Linear Regression among others, have been widely applied, aiming to detect key features of patients' conditions, health risks, as well as diseases' progression after treatment, exploiting information that derives from various complex medical datasets. ML algorithms for disease detection are algorithmic models that analyze medical data to find signs of diseases before they become severe. ML disease detection are used for imaging analysis, signal processing, and identifying multiple pathologies. In disease detection can also evaluate genetic markers for mutations and analyze biomarkers.

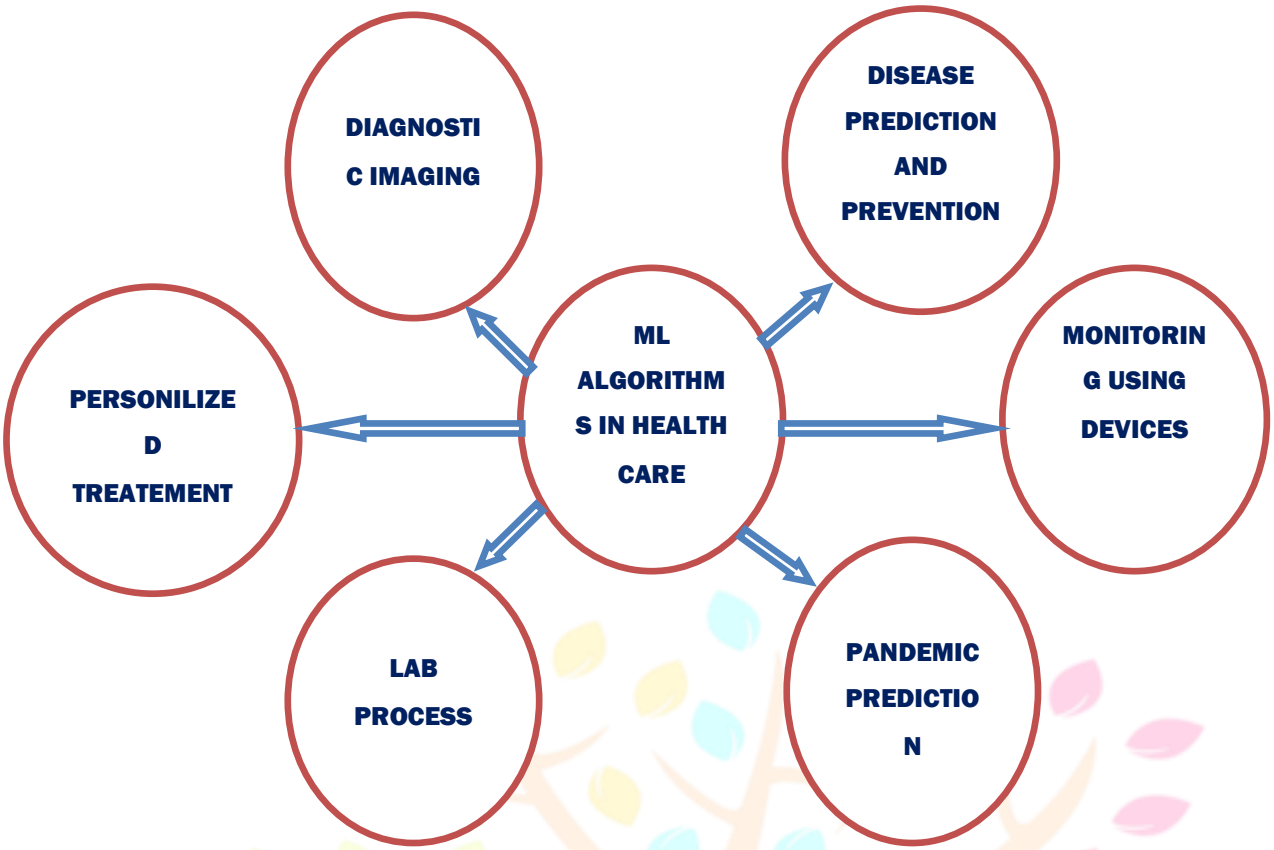

Fig 1.Machine Learning Algorithms in Health Care

IV. METHODS AND MATERIALS:

Predictive analytics, a branch in the domain of advanced analytics, is used in predicting the future events. It analyzes the current and historical data in order to make predictions about the future. Developing an environment for producing the predictive analytics models is fundamental to the effectiveness for decision-making. Here the following table summarized the different approaches and its prediction accuracy rate.

Sl.no	Existing Method	Analytics Method	Prediction Accuracy
1.	Prediction Of Heart Disease Using Hybrid Linear Regression	Linear Regression	89.13%
2.	Prediction of Heart Disease using Multiple Linear Regression	Multiple Linear Regression	60.1%
3.	Predictive Analysis Of Heart Diseases With Machine Learning Approaches	Hybrid Linear Regression	89.13%
4.	Prediction of Heart Disease Using Decision Tree	Decision Tree	68%
5.	Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers	Naïve Bayes	86%
6.	Heart Disease Prediction System using Naïve Bayes and Jelinek-mercer smoothing	Naïve Bayes	78%

7.	Predicting Heart Diseases In Logistic Regression Of Machine Learning Algorithms	Logistic Regression	87%
8.	Logistic Regression Models in Predicting Heart Disease	Logistic Regression	84.98%
9.	Neural Network based Heart Disease Prediction	Neural Networks	84%
10.	Prediction system for heart disease using Naive Bayes and Particle Swarm Optimization	Naïve Bayes	87.9%

Table 1.1 Summarization of different approaches of Predictive Analytics

- K.Srinivas proposed Hybrid Linear Regression Model (HLRM) implemented in two phases. Initially, data preprocessing is done; missing values are imputed with KNN and simple mean imputation and next Principal Component Analysis is used to extract the most contributing attributes for the cause of disease. Second, Stochastic Gradient Descent is the linear regression used to record the probability values of dependent variables, in order to determine the relationship between the dependent and independent variables. The overall prediction accuracy of the proposed model is observed as 89.13%.[1]
- K.Polaraju proposed in his experiment that he is performed using training data set consists of 3000 instances with 13 different attributes. The dataset is divided into two parts that is 70% of the data are used for training and 30% are used for testing. Based on the experimental results it is clear that the classification accuracy of Regression algorithm is 60.1%.[2]
- Ramesh TR proves in his article that the uses of ML models are the best choice to predict and classify heart disease even if the database is more diminutive. Various performance parameters, i.e., precision, F1 score, accuracy, and recall, have been compared for the entire Machine Learning classification techniques on the UCI online heart disease data set. The accuracy of 89.13% is obtained with the hybrid Linear Regression method. [3]
- Mrs. Mehdi Khundmir Iliyas is successfully developed a system using Decision tree using J48 Algorithm the develop system pull out the knowledge from historical database made by medical practitioner. In first step training is given to the system then prediction take place. The accuracy of this system is 68% true, and 32% false with the size of tree is 25 and leaves are 15 it's the smallest one as compare to other method. As this approach gives 68% accuracy only. [4]
- V Sai Krishna Reddy and his team used two algorithms Decision Tree (DT), Naïve Bayes (NB), and implemented them on the Heart Failure dataset and, recorded the results. They have analyzed 2 classifiers in-depth namely Gaussian Naïve Bayes, Decision tree and they recorded the best accuracies for the considered dataset which are 86.0% and 82.0% respectively. Recognizing the problem with less time helps to save lives and precautions to be taken for avoiding heart disease with help of ML Techniques. They conclude that the Gaussian Naïve Bayes algorithm performs well on the Heart Failure Dataset with an accuracy of 86.0%. [5]
- Ms.Rupali developed Heart Disease Prediction System using both Naive Bayesian Classification and Jelinek-mercer smoothing technique. Jelinek-mercer smoothing technique is the more effective than naive bayes to predict patients with heart disease. This model could answer complex queries, each with its own strength with ease of model interpretation and an easy access to detailed information and accuracy. The predicted rate is 78% using Naïve Bayes and 86% using classification technique respectively. [6]
- According to this paper A. S. Thanuja Nishadi used the logistic regression approach. The outcome is predicted based on classification problem. The predicted accuracy of the model is 87%.[7]
- Yingjie Zhang predicts the risk of suffering from heart disease among the elderly by exploring the feasibility of using logistic regression models. The prediction accuracy of the final model is 84.98%.[8]
- Mrs. K. Uma Maheswari concluded in her research with the integration of logistic regression and neural network gives the novel approach in predicting the heart disease of an individual. The performance measures of neural network are calculated using various measures such as accuracy, specificity and sensitivity. The accuracy obtained by the neural network is 84%.[9]
- Uma N Dulhare proposed the predictive model with Naïve Bayes accuracy is 79.12% and result recorded with Naive Bayes+PSO is 87.91%. This model improved the accuracy 87.9%.[10]

V. CONCLUSION:

This study has addressed the use of predictive analytics of heart diseases using some machine learning approaches combine with data mining techniques. It is to determine that the prediction performance of each algorithm and apply the proposed system for the area it needed. Also we can use more relevant feature selection methods to improve the accurate performance of algorithms. In conclusion, as identified through this literature survey, believe only a marginal success is achieved in the creation of predictive model for heart disease patients and hence there is a need for combinational and more complex models to increase the accuracy of the predicting the early onset of heart disease.

REFERENCES:

- *[1] Prediction Of Heart Disease Using Hybrid Linear Regression*
- *K. Srinivas1, B. Kavitha Rani2, M. Vara Prasad Rao3, Raj Kumar Patra4, G.Madhukar5, A. Mahendar6 , European Journal of Molecular & Clinical Medicine ISSN 2515-8260 Volume 07, Issue 05, 2020*
- *[2] Prediction of Heart Disease using Multiple Linear Regression Model 1K.Polaraju, 2D.Durga Prasad 1M.Tech Scholar, 2Assistant Professor © 2017 IJEDR | Volume 5, Issue 4 | ISSN: 2321-9939*
- *[3] Predictive Analysis Of Heart Diseases With Machine Learning Approaches Ramesh TR, Umesh Kumar Lilhore, Poongodi M, Sarita Simaiya, Amandeep Kaur, Mounir Hamdi, Malaysian Journal of Computer Science. Data Analytics, Healthcare Systems and FinTech Special Issue 1/2022*
- *[4] Prediction of Heart Disease Using Decision Tree by Mrs. Mehdi Khundmir Iliyas*
- *Faculty, MCA Department ,Allana Institute of Management Sciences Pune.*
- *[5] Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers V Sai Krishna Reddy, P Meghana, N V Subba Reddy, B Ashwath Rao, AICECS 2021*
- *Journal of Physics: Conference Series 10.1088/1742-6596/2161/1/012015*
- *[6] Heart Disease Prediction System using Naïve Bayes and Jelinek-mercer smoothing Ms.Rupali R.Patil, International Journal of Advanced Research in Computer and Communication Engineering, ISSN (Online) : 2278-1021.*
- *[7] Predicting Heart Diseases In Logistic Regression Of Machine Learning Algorithms By Python Jupyterlab A. S. Thanuja Nishadi University of Colombo, Faculty of Graduate Studies, Sri Lanka, International Journal of Advanced Research and Publications ISSN: 2456-9992, Volume 3 Issue 8, August 2019.*
- *[8] Logistic Regression Models in Predicting Heart Disease by Yingjie Zhang, Lijuan Diao, Linlin Ma, Hebei University of Technology, Tianjin, China , North China Institue Aerospace Engineering, Langfang China,Chinese Research Academy of Environmental Sciences, Beijing China. CCSIE 2020 Journal of Physics: Conference Series **1769** (2021) 012024 IOP Publishing doi:10.1088/1742-6596/1769/1/012024*
- *[9] Neural Network based Heart Disease Prediction by Mrs. K. Uma Maheswari, Assistant Professor, Department of Information Technology, Anna University (BIT Campus), Tiruchirappalli-24. International Journal of Engineering Research & Technology (IJERT)*
- *ISSN: 2278-0181, RTICCT - 2017 Conference Proceedings*
- *[10] Prediction system for heart disease using Naive Bayes and Particle Swarm Optimization Article · January 2018 DOI: 10.4066/biomedicalresearch.29-18-620,Uma N Dulhare Muffakham Jah College of Engineering and Technology, Banjara Hills, Hyderabad, India. ISSN 0970-938X*